Discovering Naturally Occurring Antifreeze Peptides from Microbiome by Integrating Protein Language Models and Molecular Dynamics Simulation

Ibrahim A. Imam, [a] Trevor Morey, [a,b] Yuexu Jiang, [a,c] Dong Xu, [c] and Qing Shao* [a]

[a] Department of Chemical and Materials Engineering, University of Kentucky, Kentucky 40506, USA

[b] Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA.

[c] Department of Electrical Engineering and Computer Science and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA.

Table S1 Summary of AFP and non-AFP sequence counts

Sequence Class	Random Forest Dataset	PLM Datasets	Total
AFP	1100	6506	7606
Non-AFP	1100	65060	66160
Total	2200	71566	73766

Table S2 The details of initial system setup

System	Dimension (nm×nm×nm)	System size
Peptide #1	9.4 × 7.4 × 6.0	54630
Peptide #2	9.4 × 7.4 × 6.0	54658
Peptide #3	9.4 × 7.4 × 6.0	54719
Peptide #4	9.4 × 7.4 × 6.0	54671
Peptide #5	9.4 × 7.4 × 6.0	54748
Peptide #6	9.4 × 7.4 × 6.0	54747

Tables S3 The 5 vector E-descriptors of each amino acid proposed by Venkatarajan & Braun

Amino Acid	E1	E2	E3	E4	E5
A	0.008	0.134	-0.475	-0.039	0.181
C	-0.132	0.174	0.070	0.565	-0.374
D	0.303	-0.057	-0.014	0.225	0.156
E	0.221	-0.280	-0.315	0.157	0.303
F	-0.329	-0.023	0.072	-0.002	0.208
G	0.218	0.562	-0.024	0.018	0.106
H	0.023	-0.177	0.041	0.280	-0.021
I	-0.353	0.071	-0.088	-0.195	-0.107
K	0.243	-0.339	-0.044	-0.325	-0.027
L	-0.267	0.018	-0.265	-0.274	0.206
M	-0.239	-0.141	-0.155	0.321	0.077
N	0.255	0.038	0.117	0.118	-0.055
P	0.173	0.286	0.407	-0.215	0.384
Q	0.149	-0.184	-0.030	0.035	-0.112
R	0.171	-0.361	0.107	-0.258	-0.364
S	0.199	0.238	-0.015	-0.068	-0.196
T	0.068	0.147	-0.015	-0.132	-0.274
V	-0.274	0.136	-0.187	-0.196	-0.299
W	-0.296	-0.186	0.389	0.083	0.297
Y	-0.141	-0.057	0.425	-0.096	-0.091

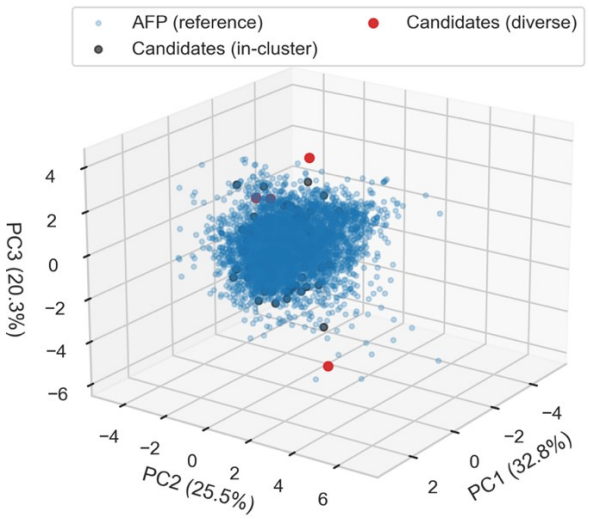
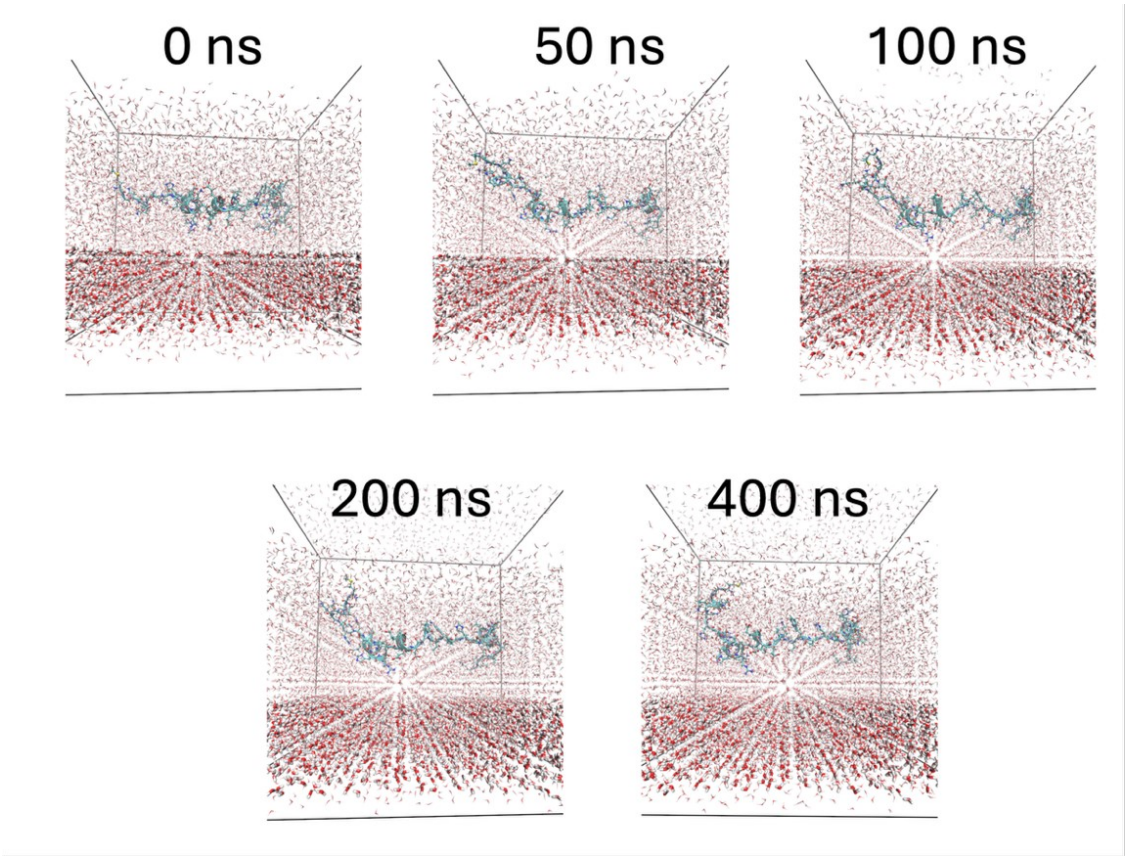


Figure S1: The 3D PCA visualization of AFP predicted peptides overlapped with AFP from the training data.

A



B

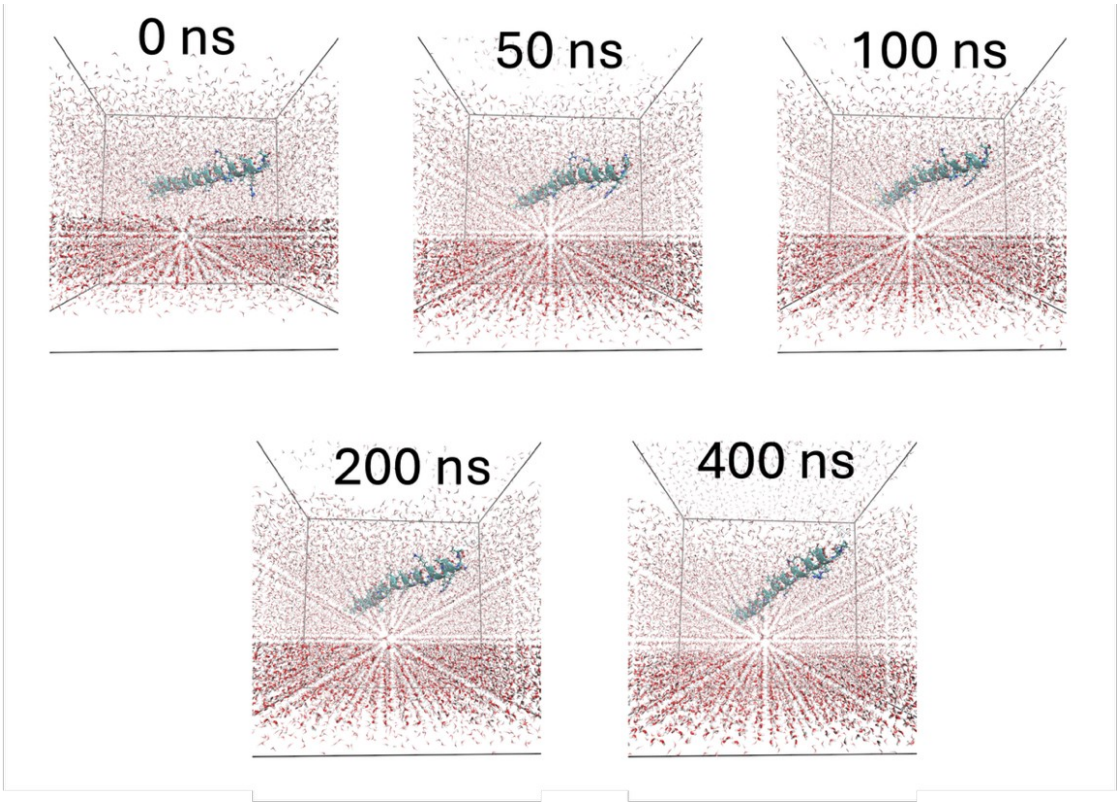


Figure S2: Figure S6. Representative VMD snapshots of the ice–water slab simulation for peptide #5 at 0, 50, 100, 200, and 400 ns for (A) peptide #1, (B) peptide #5. The restrained basal-plane ice slab maintains an ordered lattice, while the adjacent liquid water exhibits interfacial ordering.

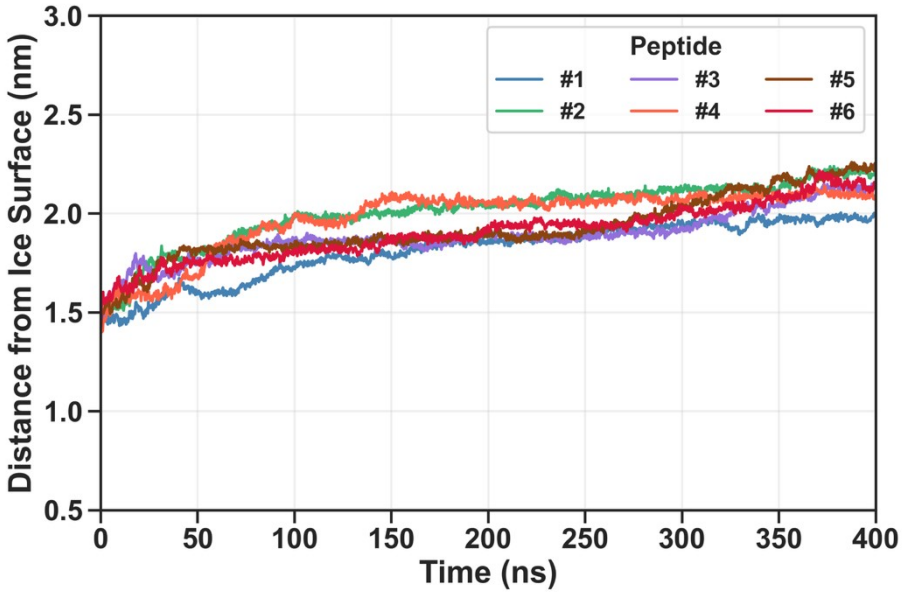


Figure S3: Peptide-ice interfacial gap (surface-to-surface separation along the slab normal, z) as a function of simulation time for the six peptide candidates in the basal-plane ice–water slab system (peptides initially placed ~1.5 nm from the ice surface).

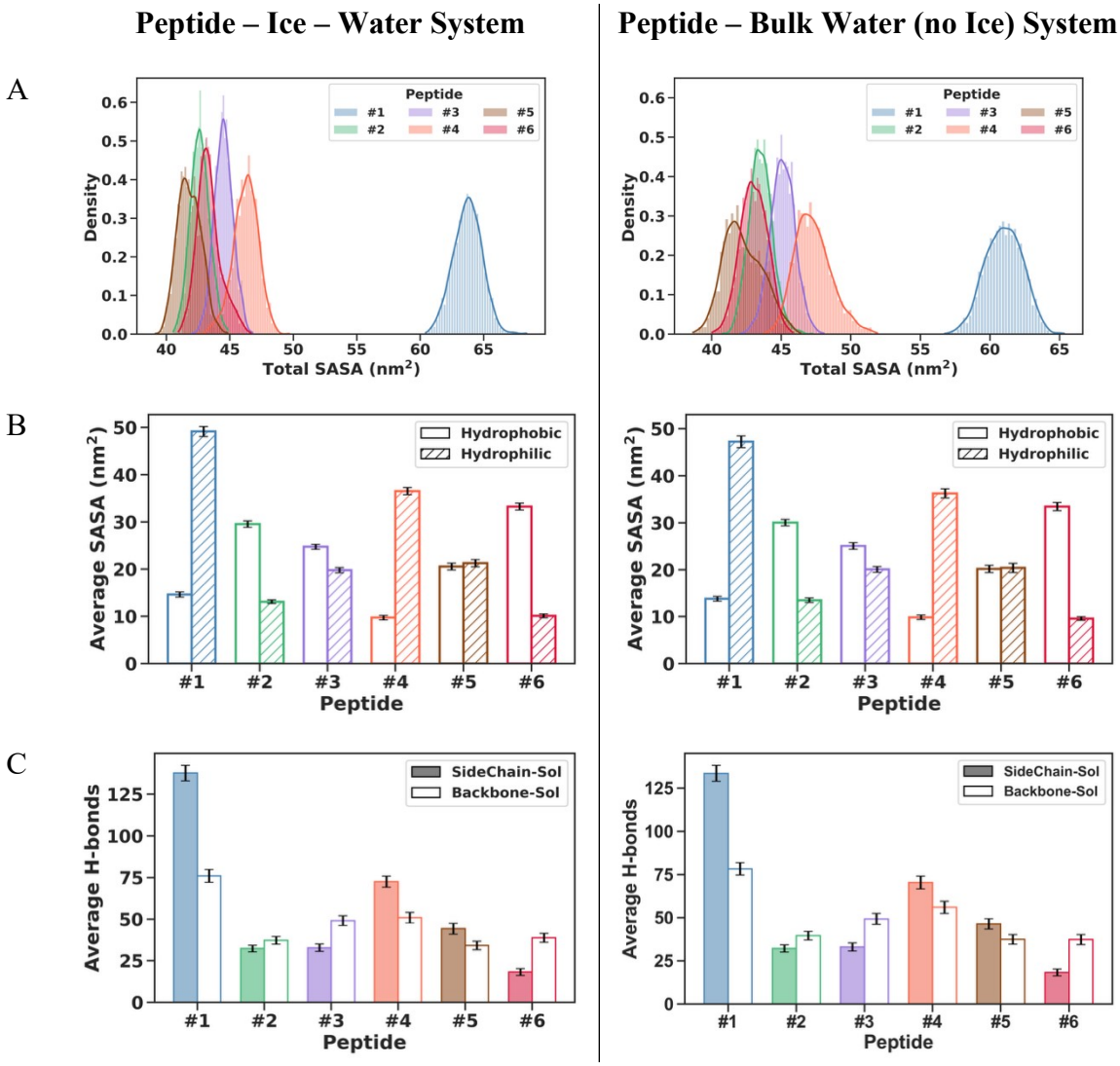


Figure S4: Comparison of key structural/solvation descriptors between peptide–ice–water slab simulations and matched bulk-water controls. Left panels: basal-plane ice–water slab; right panels: bulk water (no ice), both at 250 K for the liquid/peptide region (ice lattice restrained at 200 K in slab runs). Metrics computed over the production window ($t \geq 50$ ns). (A) Distributions of total solvent-accessible surface area (SASA). (B) Mean hydrophobic and hydrophilic SASA components. (C) Mean peptide–water hydrogen-bond counts partitioned into backbone–water and sidechain–water contributions.